

## LEVELS OF MIND

Specialization and Cooperation in the Brain

A Useful Abstraction: The Pyramid of Mind Revisited

Foundations: The Implicit Mind

    Associative Learning: From Conditioning to Categories

        Simple Conditioning

        Generalization, Discrimination, and Concept Formation

        Trees, Webs, and Other Abstractions

How Implicit Processing Really Works (Maybe)

    Neural Networks and their Attractions

    Mapping Inner Space

        Partitions: How Categories and Hierarchies Emerge

        Taking Averages: Statistical Thinking

        Truth and Beauty: Forming Abstract Ideas

        Comparing Comparisons: Analogical Reasoning

Coming Into the Light: Consciousness

    Defining Consciousness

    The Easier Questions: Characteristics of Consciousness

        The Theater Metaphor

        The Neurology of Consciousness

            Structures

            Processes

    Binding Consciousness

    A State of the Organism Meeting: A Purpose, Perhaps?

    Consciousness in other Animals: Before Us, Besides Us

        When Did it Evolve?

        What is it Like for Them?

Further Elaborations: Beyond Trial and Error

    Explicit Thinking: Virtual Test Runs

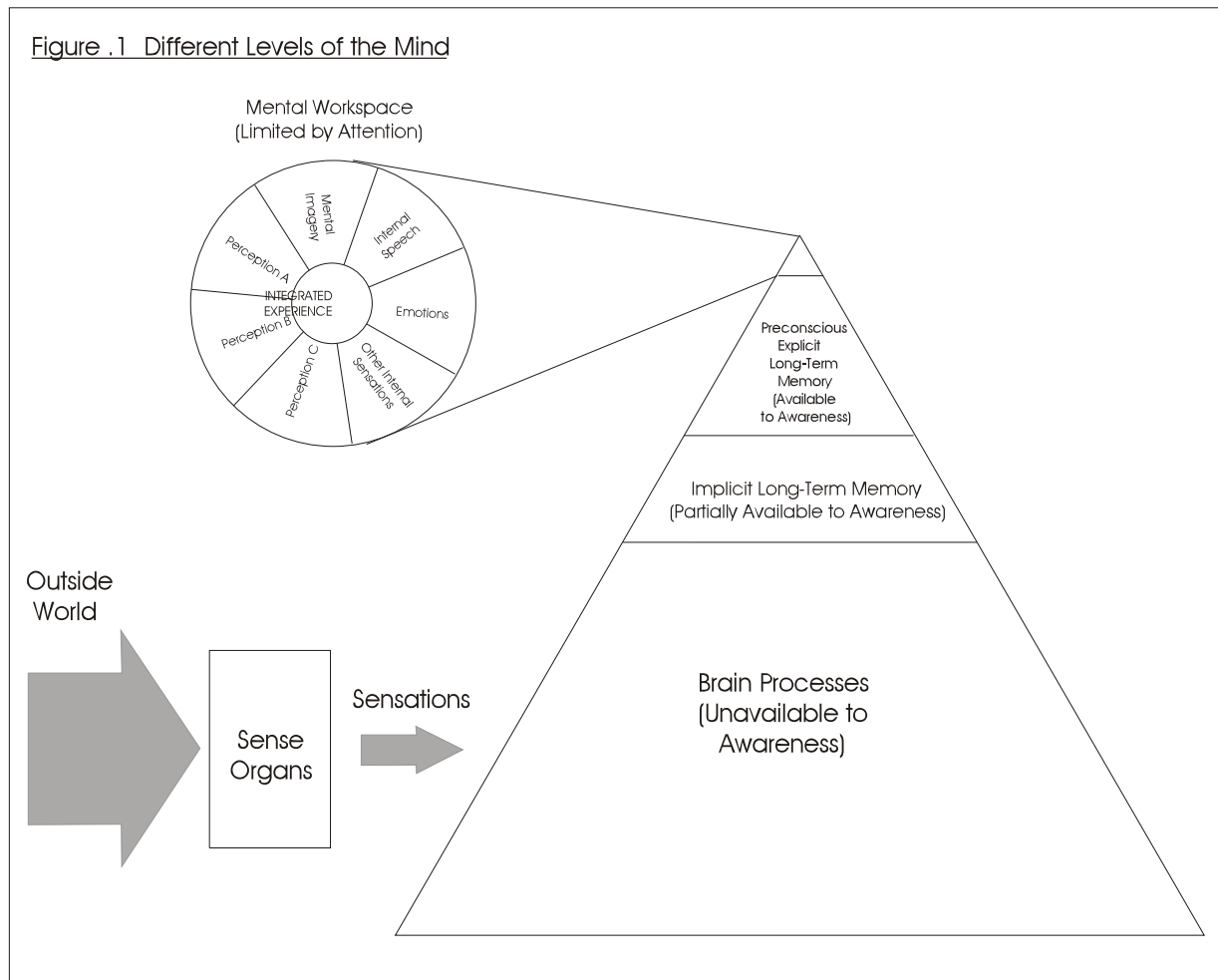
    Learning from Others: The Beginning of Culture

## **SPECIALIZATION AND COOPERATION IN THE BRAIN**

The sketch of visual processing in the last chapter gives a good idea of some general features of the brain's functioning. Different areas of the cortex, and of the brain as a whole, do different, specialized jobs. These areas are combined into complex pathways which perform complex tasks. This offers a solution to a longstanding debate about the brain. For many years, scientists debated whether brain function is localized, or whether it is distributed throughout the brain. The neurologist John Hughlings Jackson offered a solution to the dispute, by distinguishing between functions, with a small "f", and Functions, with a large "F". Small-f functions, such as line orientation recognition, or even face recognition, are most definitely localized. Large-F Functions, however, may not be so localized. They may use many specialized processors from across the brain. Let's say you see John, the husband of your friend Beth, on the street. Visual areas go to work, processing his image from the retina to the primary visual area, and then along various pathways. The "what" area of the brain, or more specifically the "who" area of face recognition, helps identify who you are seeing. Temporal lobe areas help make the association that this is Beth's husband. You also remember, with a flush of emotion rising from the limbic system, that Beth's mother is in the hospital. Frontal lobe areas concerned with planning actions activate the language areas on the left side of the brain, which help you say "Hi, John. How's Beth's mom doing?" That complex, spur of the moment process is a Function, and it needs the cooperation of many localized functions to work.

### **A USEFUL ABSTRACTION: THE PYRAMID OF MIND REVISITED**

We have moved from the basic mechanics of vision to a broader perspective on how the brain works, and indeed, how the mind works. But there are many questions remaining, which you may be asking right about now. For one thing, we haven't come up with an alternative to grandmother cells. Visual information does not converge onto cells much more complicated than those for face recognition. In fact, visual information *diverges* after it leaves the primary visual area. Where, then, does all the information "come together" into a unified, conscious perception? And speaking of consciousness, what about all that stuff from Chapter two about different levels



of the mind? I spoke there of a high-capacity, but unconscious, realm of memory and thinking called *implicit processing*, as opposed to conscious, but limited capacity, *explicit processing*. What about the hierarchical structure of concepts in long-term memory, where chihuahuas and schnauzers are filed under the heading of dogs, which in turn is filed under the heading of animals? What about the distinction between long-term and short-term memory, or the idea of selective attention?

Figure .1 reproduces the image from the second chapter of Volume I which I used to show the relationship between some of these processes. Unconscious and implicit types of processing are at the bottom of the triangle, to show their huge capacity. Farther up is the vast store of things in long-term memory. These are *available* to explicit, conscious awareness, but most of them are not in explicit awareness at any one time. The window of awareness, limited by the bounds of attentional capacity, is shown as the magnified portion of the very tip-top of the

triangle. Now we can begin to see how these various functions fit in with the actual architecture of the brain. Roughly speaking, just as the top parts of the brain are the most recent in evolutionary history, so are the top parts of the pyramid of functions. But one doesn't directly map onto the other. One of the main lessons one finds in examining how the brain works is that processing diagrams like Figure are abstractions—useful abstractions, but abstractions nonetheless. They gloss over many of the distinctions that show up when you look at the way the brain really works. And that is one of the reasons they are useful. The brain is enormously complex; perhaps the most complex object known to humankind. Even if we did fully understand the way it functions, which we don't, the full description would be too complicated to be useful to people trying to understand the mind well enough to make the most of theirs. Hence the usefulness of abstractions.

A very good example of a useful abstraction is memory—the retention of information over time; the *product* formed by the *process* of learning. The idea of memory has permeable boundaries. For one thing, there are many types of memory. We can divide memory into long-term and short-term, or implicit and explicit types. These divisions have finer subdivisions, as well. Explicit memory can be concerned with facts, events, or concepts, each of which is processed in a different way, in a different area of the brain. Implicit memory includes complex knowledge or skills such as automatically conjugating verbs or riding a bicycle, as well as simple things such as jerking one's hand back from a doorknob that tends to shock you. Not surprisingly, while some parts of the brain are more concerned with memory than others, there is no one place you can point to and say "That's the memory part of the brain". In fact, there is not even a single process by which memory works. As we have seen, short-term memory relies mostly on moment to moment changes in the connection patterns between neurons, mediated by the amount of neurotransmitter building up in the synapses, while long-term memory relies on more permanent changes in connection strength, such as increases in the number of axon terminals and synapses. The same vagaries are associated with other common terms such as thinking or learning. This is why I use general terms like implicit *processing* more often than more specific terms like implicit memory, learning, or thinking.

## FOUNDATIONS: THE IMPLICIT MIND

With these provisos in place, let's start at the bottom of our abstract pyramid of the mind, with implicit processing, and move up toward things like explicit processing and consciousness. We have come a long way toward grasping the basics of implicit processing in the brain. As we saw in Chapter Two, implicit processing is based on the wiring of the brain, not on consciously imagining scenarios or talking to yourself. This explains a tradeoff basic to implicit processing—because it conducts automatic processes in parallel, it has a huge capacity, but because those processes are automatic, it is relatively inflexible. For example, you can drive, scratch your head, and chew gum at the same time, while paying attention only to the conversation you are having on your cell phone, because the first three are mostly automatic. You could never have talked on the cell phone while you were learning to drive, because you had not yet made the various skills involved in driving automatic. But letting your driving go on autopilot has a price—you are noticing less about your surroundings, because your attention is occupied with your conversation, and you won't be able to respond well to unusual circumstances. If you need to merge suddenly with traffic, because you didn't notice that the road is becoming one lane, you had better switch your attention from the cell phone to your driving.

So, implicit processes are simply those that let all the brain's little specialized modules do their thing without the conscious mind becoming aware of them (actually, the conscious mind is often aware of the results of processing, just not the processing itself). The brain simply does what it is wired to do without reflecting on what it is doing. Sometimes this wiring is set from birth, and sometimes it is adjusted through learning. But this is a shallow characterization of implicit processing, which doesn't answer some fundamental questions. For example, how, exactly, does the wiring of the brain handle complex tasks such as driving or conjugating verbs? When learning is involved, how does the brain wire itself, changing its own neural connections in an appropriate way? We don't completely know the answers to these questions, but we do have some plausible ideas. Let's explore those ideas by looking at one of the foundations of implicit processing: associative learning.

## **ASSOCIATIVE LEARNING: FROM CONDITIONING TO CONCEPTS**

As we discussed in Chapter Two, associative learning is based on making connections between things or events that seem to go together. This may mean that they co-occur, as when cowboys tend to appear with cows, when the sound of a bell precedes the arrival of a treat, or when the act of pushing on a door causes it to open. Things may also go together because they are similar—stronger connections are made between televisions and radios than televisions and telephones, because televisions and radios perform a similar function. Now that we have explored the ways of neurons, we can look at how they support associative learning. As the saying goes, “Neurons that fire together, wire together.” Neurons that tend to fire at the same time become more strongly connected, while those that don’t become more weakly connected. When we repeatedly see cowboys around cows, the neurons that code for each become linked together. Because we seldom see cowboys around sailboats, that connection is never made. Linking things that co-occur works in the same way. If the neurons that detect the sound of a bell fire around the same time as those that detect the arrival of a treat, they will wire together<sup>1</sup>.

## SIMPLE CONDITIONING

In evolutionary terms, associative learning is the most ancient type of learning. Even habituation and sensitization are forms of associative learning, because they involve changing the connection strengths between neurons. Animals that are literally brainless are capable of this sort of learning, which may only involve one or two synapses. But the simple principle of association turned out to have great possibilities, and came to support a whole hierarchy of increasingly sophisticated processes. A step up the scale of sophistication from habituation and sensitization is **classical conditioning**. This is the phenomenon studied by Pavlov, where a response automatically linked to one stimulus is transferred to another, co-occurring, stimulus. Pavlov’s dogs automatically salivated when they saw and smelled food. When a bell was rung before the food was presented, they learned to salivate when they heard the bell, even when no food was present. Classical conditioning can easily be explained in terms of strengthening connections

---

<sup>1</sup>Conditioning experts will find that generalization woefully simplistic, and it is. This entire chapter is, because it attempts a lightning survey of how the brain works. A full treatment could fill a small library.

between neurons. Those neurons associated with the bell become connected to those associated with food and salivation, so that the sound of the bell causes salivation.

A little more sophisticated than classical conditioning is **operant conditioning**, where an animal learns to perform an action that is associated with an outcome. A classic situation is a pigeon placed in a box with a lever. Pressing the lever causes a food pellet to drop into the box. When the pigeon accidentally presses the lever, it is surprised to see a tasty treat drop into its cage. This doesn't have to happen very many times before the pigeon learns to push the lever on purpose, in order to get a snack. Instead of a link between stimuli and responses, this is a link between an action and an outcome. A positive outcome, like a treat, will encourage the behavior that leads to it, while a negative outcome, like an electric shock, will suppress the behavior that leads to it. It is yet another example of feedback in nature. As with other forms of associative learning, it is based on strengthening the connections between groups of neurons (some excitatory, and some inhibitory).

## GENERALIZATION, DISCRIMINATION, AND CONCEPT FORMATION

Associative learning can also explain somewhat more complex cognitive functions, such as how humans and other animals improve their abilities of generalization and discrimination; of lumping and splitting the things in their environment in meaningful ways. Of course, many animals, and even many non-animals, can parse their environment to some extent without learning. A bacterium splits its environments into coarse divisions such as “Good (proceed)” and “Bad (retreat)”. Of course this is a purely automatic division. It is not conscious (presumably), nor is it learned. It doesn't even require a nervous system. Even an organism as relatively sophisticated as a toad seems to rely on hardwired parsing of the world. Anything small and moving is generalized as “edible bug”. The fact that toads will eat small, moving steel balls illustrates the limitations of hard-wiring. If small boys who liked being mean to toads were common enough to exert a selection pressure, toads might become a bit more discriminating. They might gain the ability to learn finer divisions, such as that between smooth, round, inedible balls and six-legged, elongated, juicy bugs. Generalizations might also be useful. Learning that a motionless thing with six legs is also a bug, and likely edible, could be useful. Alas, such things

seem beyond the abilities of toads. But more sophisticated animals, such as birds and mammals, are quite capable of learning to generalize and discriminate.

Once an animal is able to learn to generalize and discriminate, to arrange things according to their similarity, and to link and separate things according to co-occurrence, it has the basic tools for forming the networks and hierarchies of concepts that we discussed earlier. My dog, for example, gets excited when she sees me tie my shoes, reach for my car keys, or put on my coat, because she associates all these things with getting in the car, or going for a walk, both of which she likes. She has linked these things in a web of association. Such webs are mainly based on co-occurrence. Shoes, keys, and walks don't resemble each other, but they do occur together. Many animals can also form simple hierarchical concepts. My dog can generalize across many coats—a light jacket causes the same excitement as a parka. She knows that any coat might lead to a walk. But she may not make terribly sharp discriminations. I've never noticed that she gets more excited when seeing me put on hiking boots than dress shoes, even though one is much more likely to be associated with a walk.

## TREES, WEBS, AND OTHER ABSTRACTIONS

I have slipped back into abstraction. I have been speaking of associative learning, and the many levels of function to which it gives rise, in the most general terms. One could get the impression from the discussion above that the nervous system is just one big, homogenous network that learns simply by refining its connections. The nervous system is actually quite specialized. Different parts do different things, and most of these parts are specifically wired from the beginning to do a certain job, though that wiring may become highly refined. Different types of associative learning also occur in different places in the brain and nervous system. Simple conditioning, for example, can occur entirely within the spinal cord. Some generalization and discrimination occurs in sensory areas of the brain, and some occurs in memory areas. Finally, in terms of concept formation, the nets and hierarchies are also abstractions. One shouldn't get the idea that a discrete section of a dog's brain is devoted to the idea of shoes, and another section to walks. One can't map a hierarchy of ideas across the surface of the cortex. These things are distributed across the brain, and their actual existence there is quite subtle.

This is similar to the idea of a grandmother cell. A dog, for example, doesn't have a shoe cell in its brain, or even a shoe part of the brain. When my dog sees me put on my shoes, various parts of her brain become activated to reflect that concept. Parts of the visual area devoted to shape recognize a particular shape, while others recognize the motion of putting them on. Specific parts of her brain may recognize a characteristic posture or facial expression that I adopt in these situations. All this causes memory areas to light up, pulling up general memories of past walks. Ideas like "walk", "shoe", or "grandmother" are manifested as patterns of activation that appear across many parts of the brain.

## **HOW IMPLICIT PROCESSING REALLY WORKS....MAYBE**

Now, such a pattern of activation is latent in the connection strengths between billions of neurons across the brain, so the potential is there in the wiring. But this potential isn't realized until that particular population of neurons starts to fire in that particular pattern. So, each time a concept or memory is activated, it is reconstructed. And each time is a little different. Now, the question I want to investigate is, how is that vast population of widely dispersed neurons able to assemble itself into that particular pattern when it needs to? This brings us back to the ideas of self-organization and attractors.

## **NEURAL NETWORKS AND THEIR ATTRACTIONS**

Scientists have theorized about how information processing in the brain might work by studying computer models called **neural networks**. As the name suggests, these are simulated networks of interconnected nodes, as in Figure . Oftentimes, simple networks will consist of three layers of these nodes. Let's say that we have a network for recognizing letters of the alphabet. The nodes of the bottommost layer are activated by simple lines, something like the simple cells in the visual area of the brain. These nodes are connected to a second layer of nodes, which in turn are connected to a third layer. As in connections between neurons, the connections between nodes can be excitatory or inhibitory. They can also come in a range of strengths. Some are more strongly inhibitory than others, and some are more excitatory than others. If a node in layer

one codes for a simple feature that forms part of a certain letter, it will have excitatory connections, via the middle layer of nodes, to the third layer node that codes for that letter.

If a network has its connections set correctly, it will be able to recognize letters very accurately. If an “R” is presented to the bottom layers of the network, the nodes that correspond to all the differently oriented lines that make up that letter will become active. Nodes that do not appear in the letter “R” will be inhibited. When the node corresponding to “R” lights up, the network has recognized the input.

To perform this feat, the network has to have the connections between its nodes set correctly. This means they need to be appropriately excitatory or inhibitory, at the appropriate strength. This is done by “training” the network, as follows. Initially, the connections are set at random. Then the network is shown stimuli, in this case letters. At first, of course, the network won’t be able to recognize anything, and its output will have little to do with its input. Each time, however, the actual output is compared with the desired output, and an algorithm is applied that adjusts each node in a way that brings the overall output a little closer to the desired output. After many of these training sessions, with many different letters, the system will “learn” to recognize letters very accurately. The interesting thing about this process is that the trainer does not tune each individual node by hand. All the nodes are tuned by an automatic processes. In fact, the trainer may not even know exactly how the network’s connections are set, or how they combine to produce the output they need to produce.

When a neural network’s third layer node for the letter “R” lights up, it does so because the network has assumed a particular pattern of activation. Let’s think of this in terms of the multidimensional spaces we discussed in Book One. The activation of each node can be thought of in terms of a point on along a line, a single dimension. The activation of two nodes, then, can be represented by a point on a plane, a simple XY axis. The activation of three nodes can be represented as a point in three-dimensional space, and the activation of more nodes can be represented as a point in a multidimensional, impossible to visualize, space. When a neural networks is trained, it learns to assume a pattern of activation that corresponds to a certain point in a multidimensional space, a point that corresponds to the recognition of a single letter. In other words, it develops a set of attractors, one for each letter. If we think of the activation of the network as represented by a ball in a multidimensional space, showing the network an “X” causes

the ball to roll into one place, and showing it a “T” causes it to roll into another.

The power of neural networks comes from the fact that the ball doesn’t have to start from the same place every time in order to land at the place it needs to land. This makes neural networks very flexible. For example, if you show a well trained network an R that is partly covered up, it will usually be able to find its way into the “R” attractor. The same is true if you present it with an “R” in a different font than it is used to, or a handwritten “R”. This sort of flexibility is very difficult to achieve in conventionally programmed computers, which require very precise inputs in order to make sense of things.

In many ways, then, neural networks are much more like the human mind than conventional computers are. Like us, they can make sense of imprecision. They also show a property known as *graceful degradation*. If you go in and randomly change a few of the connections between nodes in a neural network, it will still be able to perform, just not as well as before. If you tried this with a conventional computer program, changing just a couple of letters of its code, chances are that it will crash completely. The human brain also shows graceful degradation. Minor brain damage can impair the ability to perform certain tasks, but it doesn’t usually destroy it completely. Indeed, all of us are losing brain cells every day, but it seems not to affect our abilities at all, at least until old age.

Of course, computer neural networks are different from real networks of neurons in several ways. Nodes in many computer networks can take a range of activations, while real neurons fire with the same strength every time. Simple neural net programs make no real distinction between synaptic potentials and neural potentials. Of course, there is no trainer in the brain informing its networks what is expected of them, although there are probably feedback mechanisms which perform a very similar function. So, neural network programs are not a perfect model of the brain, but they do give us some very good ideas as to how the brain might do some of its jobs.

## MAPPING INNER SPACE

Since one of the themes of this book is that both nature and the human mind are arranged in hierarchies, let’s look at how the neural network model of the brain might explain our mental

hierarchies. Let's say we are thinking of a particular member of the category "dog"—a basset hound, for instance. The idea of a basset hound is represented as concurrent activation of patterns across many of the brain's specific processors. Internal imagery is a result of the visual areas of the brain lighting up in response to activation from memory areas of the brain. Shape areas become activated, in patterns representing the basset hounds rather unique shape—big nose and paws, long ears, and low slung stance. Color areas contribute color information. Motion areas may even help generate images of a characteristic gait. Temporal lobe memory areas may recall occasions when you have encountered basset hounds, or provide general impressions of their common traits. Auditory areas may generate memories of their baritone bark. All this activity combines to produce a mental impression of basset hounds. It may be some combination of internal dialogue, visual imagery of characteristic features, and memories of particular events and emotions. Different people will represent basset hounds in different ways, but everyone who has any knowledge of them will be able to represent them well enough to think about them, which is what internal representations are for.

To form this representation, each area of the brain involved has fallen into an attractor representing a certain pattern of activation. A bit more abstractly, we can think of the entire brain activity associated with the representation of basset hound as an attractor; a certain point in a very complex multidimensional space representing the relative activation of countless neurons. A similar dog, such as a beagle, will be represented by a nearby point. Very dissimilar dogs, such as standard poodles or Mexican hairless dogs, will be represented by more distant points.

Now, the entire "dog" category can be thought of as the area of the abstract mental space that encloses all the points representing particular dogs. Wolfhounds will be on one side of that space, chihuahuas on the other, but all will fall within that area. If specific categories, such as dog breeds, can be thought of as points or small areas within "concept space", then larger categories can be thought of as larger, encompassing areas within that space.

## PARTITIONS: HOW HIERARCHIES EMERGE

This way of thinking offers some insight into how the brain arranges concepts into hierarchically nested categories. The tree-like, hierarchical structure is simply a way of

representing how “concept space” is partitioned. For example, the category “animal” is a large area in mental space, encompassing many smaller areas, such as “dog”, “cat”, “hippopotamus”, “ostrich”, etc. These smaller areas, especially the more familiar ones, are in turn divided into smaller sections. We can represent this as a nested partitioning of mental space, or as a tree showing a large trunk dividing into smaller branches. The tree image is an abstraction of the partitioning of mental space, which is itself an abstraction of the extremely complex way the brain represents categories.

The brain divides the world into categories for a reason—because things in different categories need to be recognized and classified. When an animal recognizes something in its environment, it is filing it under a particular mental category. What category it is filed under, “predator” vs “prey”, “food” vs “poison”, “single” vs. “married”, makes a big difference in the way it is treated. This is perhaps the most basic function of mental categories—recognition of perceptions. If we don’t have a well-developed partitioning of our categorical spaces, we are not able to recognize the differences and similarities between things. Our ability to generalize and discriminate is impaired. When I was in high school, for example, I had very little interest in cars. Consequently, most cars looked basically the same to me. I could make coarse distinctions, between a Saab and a Buick, for example, but I couldn’t tell the difference between, say, a Honda and a Mazda. One looked just like the other to me. My *perception* of cars was vague, because my *conception* of cars was vague. When I started college, however, I had a couple of friends who were very tuned in to cars. After being around them, I began to see learn the different makes. Suddenly I could tell at a glance whether a car was a Honda or a Mazda. They actually looked different, where before they had looked just the same. Most people have had a similar experience, of learning to see distinctions between things which once seemed homogenous.

It is easy to see what is happening here in terms of the mental space idea. When we learn more about a category of things, we are making new partitions within our mental spaces. Large spaces are divided into many small spaces. My vague mental space for car was partitioned into sections- into large sections like “Japanese”, “European”, and “American”, smaller sections like “Honda”, “Mazda”, and “Toyota”, and even smaller sections like “Civic” and “Accord”. The extensive partitioning helps one make discriminations. Whereas before I only had a vague slot for “car” to file things into, now I had very precise divisions like “Honda Civic”. The partitioning

also helps with generalization, because the space is partitioned at many levels. At the same time I was forming small boxes, like “Honda Civic”, I was also developing larger boxes, like “Honda”. This is why I could now see that two different kinds of car were both Hondas, because I could detect certain common features, sometimes without even knowing I was detecting them.

As I said before, all animals that can detect differences in their world have some simple means of categorization. With most animals, these categories are more or less hard-wired. In sophisticated animals like many birds and mammals, however, categories can be refined with experience. This allows them to learn to see, hear, and even smell precise differences in things that are important to them. But categories can do more than just aid recognition skills. In very sophisticated animals like humans, they are also an important foundation for thinking. The mental space idea helps illuminate how.

### *TAKING AVERAGES*

One thing that it helps explain is statistical learning and thinking. If we see many dogs in our lifetime, we will develop a richly populated and partitioned mental space devoted to dogs, with many areas representing different types. Now, what if someone asked us to describe an *average dog*? We would form a mental image of a sort of generic dog, with characteristics that describe most, but not all, dogs. It would probably be medium sized, with a long, straight tail, medium-length ears, nose and hair, and a uniform coat color. I imagine a dog resembling a Labrador or Golden Retriever, with a brown, slightly scruffy coat. I can't remember a particular time when I have seen such a dog, and I can't think of such a breed, but it is easy to imagine one. And doing so is a rather impressive task. We have to think of the main features of dogs, decide the average appearance of each one, and combine them all together into a mental image of a dog we have never seen before. Do we go through a mental checklist, bringing up each feature one at a time, and deciding what is most common? Probably not. This would take too long. What is more likely is that we take an average of all features simultaneously, by finding a point in the very middle of the multidimensional mental “dog space”. That point will automatically represent an average of all the features that dogs tend to have. Of course, this process is mostly implicit. The average person doesn't know how she thinks of a typical dog, even though she is quite able to do

so. Even the concept of this average dog is mostly unconscious. It is a very complex neural pattern, represented by a point in an abstract mental space. Only bits of this pattern come to our conscious minds, as a sort of vague mental picture.

This ability to see average characteristics is very important, because it lets us make good guesses about what the things we encounter are most likely to be, or what they are most likely to do. Let's be a step more specific, and think of average behaviors between dog breeds (represented by a point in the middle of the mental space devoted to that breed). Imagine you are walking through a park and see a large dog running toward you. If you have much knowledge of dogs, you will be much more nervous if the dog coming your way is a German Shepherd than if it is a Golden Retriever, because you feel that German Shepherds are more likely to attack you. If you looked up the statistics, you would be absolutely right. They are more likely to attack you. This is a good example of the uses of statistical reasoning. Of course, this sort of reasoning is also prone to error and abuse. For one thing, it only tells you what is most likely, not what is certain. Many German Shepherds are quite friendly. Perhaps some Golden Retrievers are vicious, though I have never met one. So, while it is prudent to be cautious if you see a German Shepherd approaching, you can't be certain that that particular one means you any harm.

## TRUTH AND BEAUTY: FORMING ABSTRACT IDEAS

The ability to imagine an average, prototypical member of a category that doesn't quite resemble any member you have ever seen before is a profound thing. It is a simple example of the mind producing an image that does not exist in reality (though it may help clarify that reality). It represents a first step along the mind's independent path from reality, a path that says a great deal about the human condition. Accordingly, philosophers have paid close attention to our ability to make abstractions. Plato believed that the fact that we have an idea of an ideal or perfect example of a category, a Platonic Form, must mean that the ideal actually exists. In fact, he believed that the world of such forms is what is truly real, and when we imagine these perfect forms, we are making contact with this eternal, transcendent world. Many people today draw similar conclusions. I have heard several people say that there is something in the human mind that could not have just arisen by natural selection, because we have an idea of perfection. If we have never

seen perfection, how could we have such a notion?

Well, easily. We can imagine the perfect person, say, in much the same way that we imagine an average dog. We find a spot in mental space that corresponds to it. Let's say that you are looking for the perfect dog to get as a pet. There are certain things that you have in mind that you are looking for, say, non-shedding, friendly, well-behaved, athletic, likes cold climates, etc. Each of these is a dimension along which dogs vary, and each is better if the dog is farther along that dimension (at least up to a point). We simply find a point in dog space that, instead of averaging all qualities of dogs, maximizes those features that we are looking for. Voila! We have formed an image of a perfect dog, even if we have never seen such a dog. Similar processes could easily account for the common human notions of a perfect mate, a perfect society, a perfect body, and so on.

While this is an impressive trick, it is not an unexplainable one. For example, many people have considered the idea of a perfect human body. This imagined perfect body bears a specific relationship to other bodies; it optimizes all the traits that are thought to make bodies beautiful. In terms of the mental space idea, we can think of the average body occupying one point, and the perfect body occupying another. The complex relationship between these two can be imagined as a line between one and the other. The idea of perfection in this case, while it may be consciously represented by some hodgepodge of images, is really represented by the brain as particular trajectory through mental space.

This helps explain a curious, and often harmful, habit that people have. We have the ability to extend that mental trajectory far outside the range of the possible. If you look at a Barbie Doll, or a comic book superhero, they have proportions that no human could ever attain. Comic book artists are able to imagine and draw such features because they take can take the mental trajectory that connects average human bodies and beautiful or impressive, but realistic, human bodies, and extend it into superhuman realms. So it is that Superman's head grows small compared to his height, his legs grow long, his chin grows prominent, his waist shrinks, and his shoulders widen. If we extend the same trajectory even farther, the image becomes an obvious caricature, as in the cartoons of huge chested, tiny-waisted women like Jessica Rabbit in the movie *Who Framed Roger Rabbit?* Here is a case where the idea of perfection, far from being a transcendent link with the divine, can be harmful if we take it too seriously. We can imagine

people with proportions that few, or even none, of us could ever expect to attain. Those who expect to are in for trouble.

Imagining an average, perfect, or caricatured member of a category, then, is a very explainable process involving the identification of certain points and trajectories through mental space. An imagined perfect member of a category occupies a particular point in mental space in relation to the rest of the space, and thus, to the average member of the category. The path between the two points has a particular trajectory. More abstract ideas such as the notion of perfection in general, or averageness in general, arise when we see similarities between the locations of these points, and the trajectories of the paths that connect them. The path between the average society and the perfect society has a similar trajectory to the path between the average body and the perfect body, namely, it goes in the direction that optimizes good features.

#### COMPARING COMPARISONS: ANALOGICAL REASONING

When identifying an abstract quality such as perfection, what we are identifying is a relationship between relationships. This ability underlies an extremely important form of inductive processing—analogy. Here's an example of this sort of reasoning. Choose the best, most meaningful, answer to the following multiple choice analogy: Winston Churchill is to England as \_\_\_\_\_ is to the United States. A. Calvin Coolidge B. Abraham Lincoln C. W.C. Fields D. Bill Clinton. Most people familiar with American and British history will pick Abraham Lincoln. While the two resemble each other not at all physically, they were both heads of state who were seen as guiding their nations through a crisis with shows of resolve and clever words of wisdom. In Churchill's case that crisis was World War II, while in Lincoln's case it was the American Civil War.

Now, how does the brain see such a relationship? It may do so in terms of connecting points in mental space. Churchill is represented in the brain by a particular point in mental space, representing his various characteristics. English history is represented by another point. Some cognitive scientists have suggested that what we do is draw a line between England and Churchill, as in Figure . This line represents a conceptual relationship between Churchill and England. Now, the United States is also represented by a point in mental space, as are all of the choices for

answers to the analogy. If we draw a line from the United States point, *parallel to the line between England and Churchill*, it will pass closest to the point representing Lincoln. Each line represents a particular relationship. When we compare the two lines, we are seeing a relationship between two relationships—an analogy.

Analogical reasoning, as in the example above, is a rather sophisticated form of thinking, able to deal with complex, higher order relationships and similarities. But, like many complex things, it is based on simpler things—the ability to see simple similarities and relationships. Many of other people listed as choices in the question above, for example, are in some way similar to Winston Churchill, but the similarity is not as deep as that with Lincoln. The shallowest one perhaps is W.C. Fields. Fields and Churchill physically (and in some ways behaviorally) resembled each other, but they played very different roles in their societies. Theirs is a lower order, surface connection, very different from the deep connection between Churchill and Lincoln.

Nonetheless, I suspect that many people would find the comparison between Churchill and Fields rather pleasing. This illustrates an interesting point. We enjoy making connections between things. Nature seems to have insured that we find pleasure in doing so, probably because it is useful to find relationships in our environment. The comparison between Churchill and Fields seems to be especially pleasing for two reasons—it is a very close match in some ways, yet it compares figures from very different domains. The more apt the connection, and the more unexpected, the more we like it. It is as though we are rewarding ourselves for seeing an insightful connection. This makes sense. We are a creative species, which thrives by finding new ways to cope in our environment. Its no wonder that we are rewarded by seeing new connections in our environment. This brings us back to the idea of creativity. We like to make connections that are novel, but still appropriate. We like to see unity in diverse things.

### **COMING INTO THE LIGHT: CONSCIOUSNESS**

The processes we have been discussing are all mostly implicit. When we construct categories, learn to discriminate and generalize, or see abstract relationships, most of the processing is going on subconsciously. We are mostly just aware of the results. We see the similarity between Churchill and W.C. Fields first, then go back and unravel what exactly connects

the two. If these processes do work in the ways I've described, it is easy to see why they are mostly unconscious. If our brains represent an average dog as a point in a multidimensional space, that representation cannot be entirely conscious, because we can't consciously imagine a space of more than three dimensions. We certainly don't experience our concept of the average dog as such a point. Take a moment to think about your idea of an average dog. The vague impressions and images that come to mind really aren't very well developed. They don't seem up to the job of performing the complex tasks involved in abstract, categorical thought. Yet we are confident that we have a good idea of what an average dog is, and that we are able to use that idea to think with. The conscious images are probably just labels we use to record the progress of mostly unconscious processing.

Yet not all thinking is so subterranean. If you imagine how you are going to redecorate your bedroom next week, you really are imagining it; manipulating conscious images. This is a mostly explicit process. These differences in conscious access to different types of thinking bring up some big issues. The fact that most of the brain's housekeeping functions, perceptual processing, and even thinking processes are unconscious suggests that unconscious processing came before conscious processing in evolutionary history. How, then, and when, did animals first become conscious? How does consciousness work, and what purpose does it serve? Are there levels of consciousness, and if so, which ones came first? None of these questions yet have widely accepted answers. Consciousness is one of the most puzzling and controversial issues in all of science, and indeed, in all of human thought. Yet it is truly one of life's great questions, and its emergence is surely one of the great transitions in natural history.

## **DEFINING CONSCIOUSNESS**

The first order of business in discussing consciousness is to specify what, exactly, you mean by that word. Different people mean very different things, even those who devote their careers to studying it. Some are talking about *self-consciousness*; the knowledge that one is a distinct entity which remains relatively separate and constant in a changing environment. Others are talking about a sort of higher order awareness—knowing that you know. These things are not what I mean by consciousness. What I am talking about is basic awareness—the having of

sensations and experience. It is what is present when you are awake, and absent when you are in a deep dreamless sleep.

I cannot help thinking that this is the crux of consciousness. The other definitions I listed above seem to be secondary features at best. Consider self-awareness. You can imagine having sensations of your body and of your environment, without being aware that the two are separate. You would still be having sensations, even if you weren't having self-awareness. Anyone who has ever "lost themselves" in a book or a movie has had something close to this experience. With that being said, self awareness of a very visceral kind is probably a basic feature of the mind, of most creatures with a mind, because one of the tasks of the brain and mind is to maintain the integrity of that creature's body against the flow of entropy. In fact, processes dedicated to the preservation of "self" are surely much more ancient than consciousness. They don't even require nervous systems. Also, one aspect of self that is inextricable from consciousness is that consciousness is an entirely first person affair. It is only experienced from one point of view—that of the organism that is experiencing it. But here again, one can imagine being the sole experiencer of the consciousness associated with your nervous system, without being aware of being a separate self. It may not happen often, but it is imaginable.

As to the idea of consciousness being founded on "knowing that you know", the only way I can make sense of this idea is if it is meant to point out the distinction between knowing without consciousness, and knowing with consciousness. For example, imagine that a spider building a web is an absolute automaton—that it has no experience. It is no more like anything to be a spider than it is like something to be a rock (I don't know if this is actually the case or not). The spider knows how to build a web, but it doesn't experience that knowledge, or anything else, for that matter. I, on the other hand, know how to make toast, and I experience something when I do it, or when I just think about whether I can do it. I know that I know how to make toast. In this sense, knowing that you know is simply another way of saying that perception or mental activity is accompanied by experience, by consciousness. And it is not a particularly good way of saying it, because the feeling that you know something is only one of the many facets of consciousness. Besides, I think that most people who use this expression really do believe that consciousness is based on some sort of higher order reflection on one's own mind—knowing that you know. But doesn't this lead to an infinite regress? To know that you know, do you have to know that you

know that you know? And to do that, do you have to.... You know what I mean.

If you are getting the feeling that consciousness is a slippery, mysterious topic, and that people have trouble even agreeing on the basic ground rules and definitions for discussing it, you are absolutely right. Different people just seem to conceptualize consciousness in different ways, and to have trouble even understanding what people mean when they talk about other conceptions. I personally see consciousness as fundamentally a raw “having of sensations”. Some people seem to find this idea meaningless. They can’t imagine consciousness without a sense of self, or the presence of complex things like language. I am not trying to belittle these ideas, though I do disagree with them. After all, I have trouble understanding what one means when they talk about the necessity of “knowing that you know”. So, be warned. Consciousness is still defined in many different ways. I’m not sure I have ever read a book on consciousness that didn’t assume a certain point of view, and I don’t think I can write one. What follows is an attempt to be fair to different ideas, but don’t get the idea that it represents any sort of consensus among students of consciousness. About the only thing they agree on is how difficult a topic it is. Some don’t even agree on that, because they are sure they have solved it. And right now, that is the only view of consciousness that truly deserves to be scoffed at.

## **EASIER QUESTIONS: CHARACTERISTICS OF CONSCIOUSNESS**

Let’s start with some of the easier questions. What are some basic characteristics of consciousness that can be studied empirically, “from the outside”? We have already seen some of them. As we saw in an earlier chapter, consciousness has a limited capacity—it can only contain so much information at any one time, so much of its complex processing has to be sequential, as opposed to parallel. This is in sharp contrast to the unconscious processes of the brain, which have massive capacity and can work in parallel. As the workspace image suggested, consciousness provides a small space where the most pressing issues, from inside and outside the body, can be brought together and dealt with. This means that several types of things can appear in consciousness. There are sensations from the world outside or from inside the body, internally generated visual imagery, as well as internally generated bodily sensations, or sounds. Some people may have vivid internally generated tastes and smells, though mine are not very vivid.

There are various types of explicit memories, such as memories of facts and events. There is commonly a running commentary of internal speech. Everything is colored one way or another by various emotions. Of course, because consciousness is limited, only a few of these things are present at any one time. If one thing expands, others have to give.

## THE THEATER METAPHOR

The psychologist Bernard Baars, one of the most pragmatic and inclusive thinkers about consciousness, summed up his ideas in a superb book entitled *In the Theater of Consciousness: The Workspace of the Mind*. Notice that two metaphors for consciousness appear in this title—a theater, and a workspace. Both of these metaphors turn up quite often, because they are very useful. Each one is good for highlighting different points, which is probably why Baars mentioned them both in his title. The workspace metaphor, which we have already encountered, is good for pointing out that consciousness is a place for bringing together inputs from various places, to bring to bear on the issue that is currently in focus. But without caveats, the workspace image can give the impression that there is a little person behind that desk. This is the homunculus; the ultimate shuffler of and witness to the contents of consciousness. This idea has problems, the biggest of which is the infinite regress it implies. If we become conscious when the little guy in our heads witnesses the contents of consciousness, how does he become conscious? Is there a little guy in his head, too?

The theater metaphor is not so prone to such funhouse images. In Baars' theater model, consciousness is like a stage full of actors, each representing a different content of consciousness. The stage is small, so it will only hold a few actors at a time. Even on the stage, some actors are more prominent than others. One or two in the center bask in the spotlight of direct attention, while those on the edges are more dimly lit. This highlights the fact that consciousness is not an all or nothing affair. Things in the center stand out, while those at the periphery fade by degrees into the background of the unconscious. What about the audience? This is not a rehearsal, with only the homuncular director watching. The house is packed, and the audience members are different parts of the brain, watching what is happening on stage to keep abreast of the current focus of consciousness. This model illustrates some vital features of consciousness. Like a stage,

it can only be occupied by so many players at one time. But many players come and go as the play progresses. While it is a small venue, it draws from an enormous pool of talent. Not only that, but the play is witnessed by a large, unconscious audience, whose actions are influenced by what they see on stage.

As you may have noticed, this image not only illustrates the characteristics of consciousness, but it also begins to hint at the uses of consciousness. But let's put that topic toward the edge of the stage for a moment. First, let's look at another set of features of consciousness that can be studied from the outside—the brain structures and processes that seem to be involved in producing it.

## THE NEUROLOGY OF CONSCIOUSNESS

### *STRUCTURES*

Down in the brain stem is a long structure running from the upper spinal cord to the thalamus, called the *reticular formation*. This structure is involved in several very basic functions such as breathing and heartbeat. One of its functions is the control of arousal, or wakefulness. Arousal can be seen as a continuum, from deep sleep to wide awake, focused alertness. The reticular formation is the main regulator of this continuum, which makes it a vital player in consciousness. Moving up into higher regions of the brain, the thalamus is the next big player. As we have seen, the thalamus functions as a relay station between various parts of the brain and the sense organs. Signals from the eyes to the visual areas of the brain, for example, pass through the thalamus, as do signals between the cortex and the brain stem. This traffic goes both ways. For example, one set of axonal fibers carries signals from the thalamus to the cortex, and another set carries signals back again. Inside the thalamus is a set of structures called the intralaminar nuclei. These seem to be absolutely critical for consciousness. Damage to the intralaminar nuclei in both thalami result in permanent coma.

You may be forming a hypothesis about now. If consciousness is a limited-capacity workspace or theater that draws on vast, unconscious resources, and the thalamus, particularly the intralaminar nuclei, are small areas with connections to the rest of the brain, does that mean

that the thalamus is the conscious part of the brain? Well, probably not. Though we are not entirely sure how the wiring of consciousness works, its limited capacity is probably not a result of signals coming together in one location. There is probably not a conscious part of the brain. The theater of consciousness is not a physical location. It is a result of a pattern of activity across the brain. While the intralaminar nuclei and the reticular formation are required for consciousness, they are not the sites of consciousness.

## PROCESSES

Perhaps we should take a closer look at this “pattern of activity” that is consciousness. If electrodes are placed around the head, they will always detect brain activity, even during deep, dreamless sleep, because the brain is always at work regulating the body. Neurons across the brain fire continually. However, the pattern of this neural activity changes dramatically depending on the state of arousal. Oftentimes, it arranges itself so that a pattern of oscillations can be discerned against the background noise. During alert waking consciousness, for example, neural activity shows distinct waves, cresting and troughing evenly at about 40 cycles per second. Brain waves near this frequency seem to be associated with consciousness. As sleep sets in, and consciousness slips away, they become slower and more erratic. At first, during those moments when we are slipping into sleep, the waves enter a spiky state called theta activity (Figure). Consciousness is still present here, but it is incoherent. I often realize that I am falling asleep when I have a thought, and then think “That didn’t make sense at all!” During the next phases of sleep, consciousness seems to vanish. People who are awakened during these states usually, but not always, report that they were not dreaming. This is very different from REM, or Rapid Eye Movement, sleep. People awakened during REM sleep almost always report that they were dreaming. Interestingly, brain waves during REM sleep are theta waves, similar to those during the transition period into sleep.

Consciousness, then, seems to be highly correlated with brain waves. Waking consciousness is associated with brainwaves of around 40 hz. Consciousness becomes incoherent when brain waves become slower and more erratic, as in the transition to sleep or REM sleep, and disappears as they become slower and more erratic still. What seems to be going on here is that

brain waves are produced by signals passing back and forth between the thalamus and other parts of the brain, particularly the cortex. How quickly and coherently these signals pass is controlled mainly by the reticular formation. When they approach 40 hz, consciousness appears and gels.

## **BINDING CONSCIOUSNESS**

This association between brain wave frequency and consciousness has led some scientists to propose that consciousness is based on the binding together of the activity of many neurons, so that they are locked together in coherent pattern of firing. This is a proposed solution to the binding problem. Instead of relying on neural signals converging on one spot, such as the intralaminar nuclei or a “grandmother cell”, neurons are bound together by firing together. A particular group of neurons, in a particular pattern of excitation, firing at a particular frequency, results in a conscious experience. Let’s say that there is a knock at your door, and you open it to find your grandmother. Millions of neurons, from basic feature detectors to complex detectors of motion, shape, tone of voice and so on, become active in various parts of the brain. As all the excitation and inhibition works itself out, some neurons become more active, while others become less active. Eventually, a certain set of neurons is left active in a certain pattern of excitation. When that set of neurons, which may be scattered across the brain, begin to resonate at 40 Hz, we become conscious of a certain perception—Granny!.

Earlier, I suggested that multiple neurons assume certain patterns by falling into attractors, which are latent in the pattern of connections between neurons, either as a result of hard-wiring or learning. If the 40 Hz binding idea of consciousness is correct, then this is not a static attractor, but a dynamic one. The neurons all begin to fire coherently at a particular, cyclical rate. So, neurons across the brain are not just bound together in the sense that they are physically connected and causing each other to fire. They are also bound together in that they are firing in a coherent temporal pattern.

This gives us a new way of thinking about the limited theater of consciousness. Those things that take center stage do so not simply because signals have reached a certain part of the brain (though certain parts of the brain probably do have to be involved) but because they are based on that set of neurons that have become most active, by mutually exciting each other into

the most coherent pattern. The limited theater is a result of the fact that only the most highly activated and coherent patterns of neural activity (in those parts of the brain that participate in consciousness) become fully conscious. This explains why those ideas lurking at the edges of consciousness are rather vague and undeveloped—they are based on patterns of neural activity that have become partly, but not fully, activated. If they do become fully activated, they will move to center stage.

If consciousness is produced by neurons firing together in coherent patterns—falling into cyclic attractors of activation—then we can imagine what is happening as we lose consciousness—the patterns of neural activity become less coherent and well-defined. When we are awake and focused, consciousness is produced by very tight patterns of activation. Let's say that you are sitting down, thinking hard about a problem. "What are we going to do about granny's heart condition?" The neurons involved in thinking about this problem are highly activated, while others are hardly activated at all. If we think in point attractor terms, the active neurons have fallen into a very deep, narrow basin of attraction corresponding to a focused state of consciousness. If you get tired, you lose your focus. The tightly bound patterns become looser. Your mind starts to wander, and other thoughts displace those about granny. Your neural activity is now in a shallower, wider basin of attraction, less focused on one idea. Interestingly, this state is where you are most likely to have creative ideas, because one pattern of activity is not so dominant that it inhibits others.

Now imagine that you are getting sleepy. Neural activity becomes even less coherent, and your brain waves slow down and become spiky. Some patterns still become active enough to be conscious, but they are so diffuse as to be poorly developed. Patterns representing granny lose their integrity, and start shifting into other patterns. Maybe your grandmother looked a little like Edith, from the TV show *All in the Family*, though you never noticed it. You may start having bizarre thoughts. "When was granny married to Archie Bunker? Was that before she met Grandpa?" After a while, neural patterns become so diffuse as to be completely incoherent, and thus, unconscious. You have fallen into a deep, dreamless sleep.

Now, this explanation of the processes behind consciousness sounds quite plausible (at least it does to me). But take it with a few grains of salt. It has not been proven, and it has commanded no consensus among investigators of consciousness. It is oversimplified, and full of

explanatory gaps. For example, not all 40 hz brainwaves are associated with consciousness, and sometimes people report that they were dreaming during the normally dreamless phases of sleep. At this early stage, though, it's as good as any other theory, and better than most. If it is on the right track, it suggests that basic consciousness—raw awareness—is a system built into the hardware of the brain by evolution, and is not a function of culture, language, or self-awareness (though these things may add new dimensions to consciousness).

One thing it does explain very well is the way that the brain is always striving to interpret what it perceives, even if that interpretation is wrong. What appears in consciousness is not so much a direct mirror of the world outside as a best guess. For example, once I was stopped at a stoplight in Denver, and was startled to see that I was crossing Xenophobia street. “Pretty conservative part of town” I was thinking, when I looked back to make sure I had seen correctly. I hadn't. It was Xenobia street, not Xenophobia street. But what I had seen—what had appeared in consciousness—was “Xenophobia”. I had never seen the word Xenobia, so my brain found the nearest familiar word. My neurons had never developed a basin of attraction for Xenobia, so they found the nearest one. What comes to consciousness is based on the pattern of neurons that is most strongly activated, and those patterns are the ones that are the best fit with the data from the outside world. Usually this works just fine, unless we are encounter something unfamiliar, but close to something familiar. In these situations, we literally hear and see things which aren't there.

Let's move up a level of abstraction, and return to the theater metaphor of consciousness. Now we can add a new twist to this image. This is not your traditional theater, where the actors wait for their cue before coming on stage. In the theater of consciousness, the actors—patterns of neural activity—are constantly struggling for the spotlight—the high level of activation that makes them the focus of consciousness. Actually, perhaps it would be more accurate to say that various groups of actors are struggling over what they will be performing. The ones that win will be those who are most activated, and who can put together the most coherent performance. Let's say you are driving along, daydreaming. Your mental theater is full of actors singing a relaxed number about your weekend plans. In the distance appears a large, four-legged form with antlers. The stage is invaded by a new set of actors, singing a frantic song- “Careful! Deer! Watch out! Brake, Brake!” The leaders of this invasion had been activated by shape detecting areas of the

visual system. As you come closer, though, other areas become active and send their own actors. Depth detectors find that the deer appears two dimensional, while color detectors notice that it is a uniform black color. There are letters along its side, which language areas decipher- “Watch for deer”. The actors singing about deer and caution lose their momentum. New actors arrive, and a new song emerges, “Just a caution sign, no problem”. The audience is glad to see this number, but gets the point quickly. Soon the daydream actors are back, and they strike up again.

This sort of image hints at the purposes of consciousness. Perhaps consciousness is like a forum, where signals from different areas of the brain come together to try to form a unified impression of what is going on—in the outside world or inside the body—and what to do about it. It is a way of bringing together information from various specialized areas, and letting them compare notes in an ongoing, constantly updated “state of the organism” meeting. In this sense, the best metaphor is not a workspace, or a theater, but a town meeting, where various citizens and groups are constantly trying to hammer out what the pressing issues are, and decide the best way of dealing with them. Sometimes they cooperate and take turns, sometimes they shout each other down. Either way, the goal is the clarification of the organism’s overall goals and how it should go about meeting them. Like democracy, it is not a tidy process, but it usually works. What we have here is another case of a distributed, compartmentalized system developing a means of top-down control. As figure illustrates, just as brains and nervous systems evolved to help the genes oversee the functioning of the organisms, perhaps consciousness evolved as a way to oversee the functions of the brain. If the brain is the interpreter and overseer of the body, consciousness is the brain’s way of overseeing and interpreting itself. Perhaps.

Some readers may be taking issue with the idea that consciousness is an ongoing “state of the organism” report. After all, the contents of our consciousness goes far beyond where we are and how we are doing in the here and now. We spend a great deal of time dwelling on things that happened long ago, or might happen well in the future. We ponder abstract ideas, such as beauty or justice. We can think about places we may never go, or the personalities of long dead people. These are important points. But just because consciousness is capable of much more than current monitoring does not mean that is not its original purpose. Our ears do an excellent job as anchors for eyeglasses, but they evolved as organs of hearing. The wonderful things that humans can do with consciousness are likely to be elaborations on an original, much simpler function. In other

conscious animals, consciousness is likely to be much more basic. This brings us to some important questions. When did consciousness first evolve? What is consciousness like in other animals, and how have we elaborated on it?

## **CONSCIOUSNESS IN OTHER ANIMALS: BEFORE US, BESIDES US**

### WHEN DID CONSCIOUSNESS EVOLVE?

Nobody knows. There is no first known fossil of consciousness. We don't even know which animals are conscious and which aren't. I very much doubt that the worm forgives the plow, but does it feel the plow in the first place? Most people seem to think not. I am not so sure. There is just no test for detecting the presence of consciousness. Of course, we could assume that only those animals with neural features like 40 Hz brain waves, intralaminar nuclei in the thalami, and reticular formations are conscious. But then, perhaps consciousness is based on other processes or structures in other animals. Abilities don't always depend on specific structures. Fish don't have ears, but they can still hear. And some animals have some of the features associated with consciousness, but not others. Reptiles don't have a well-developed cortex or intralaminar nuclei, but they do have a reticular formation, and they seem to have various levels of arousal. Are they conscious? Is a basking lizard feeling as satisfied as it looks? Who knows? About the only thing that can be said with any certainty is that both birds and mammals have all the structures associated with consciousness, so they are likely to be conscious. Some well known thinkers will dispute this, claiming that consciousness is recent and unique to humans. This anthropocentric view is, I think, very likely to be wrong. When it is used to claim that such animals are incapable of suffering, and can be treated as inanimate objects, it is reckless, and if incorrect, horribly cruel. (I should say that most people who think animals are unconscious still think they should be given the benefit of the doubt. But not all.) Right now, we simply don't know which animals are conscious and which are not, because we don't understand consciousness well enough. Perhaps someday we will.

### WHAT IS IT LIKE FOR THEM?

Let's assume, for the sake of argument, that birds and other mammals are conscious. What is consciousness like for such creatures?. What goes through the head of, say, a bear? What actors populate the theater of its consciousness? Well, there are most likely basic feelings and emotions, such as raw hunger, fear, pain, pleasure, anger, and so on. Perhaps there are more complex emotions, such as jealousy. Dogs certainly act like they feel jealous, and I would guess bears do as well. Very complex emotions, such as righteous indignation or nostalgia, are harder to ascribe to most animals. Also likely to be present are basic perceptions—sights, smells, sounds, and so on. There is no reason to think that intelligent animals can't generate such images internally. Dogs run and bark in their sleep, which indicates that they are having vivid dreams. Smart animals like dogs and bears might very well think simple thoughts with internal imagery. I can easily imagine bears daydreaming about finding a really good honey tree, though I don't know if they really do.

Another good assumption about consciousness in animals is that its contents probably depend on the animal. Visual animals such as hawks or monkeys probably devote more of their consciousness to vision than smell. As a visual animal, I know I do. With more smell-oriented animals, such as bears or weasels, the situation may be reversed. Perhaps bears can “smell with their mind's nose”, thinking in terms of imagined smells in a way that we could barely imagine. But perhaps not.

In any case, some animals are certainly capable of mental feats that are impressive by any standard. Many birds can remember thousands of places they have stashed food for the winter. Whether those memories are conscious is not clear, but it's not a far-fetched idea. Some animals seem to be far more capable of social thinking than others. Social animals, such as wolves and prairie dogs, are probably much more capable of feeling long term animosity or affection toward others than non-social animals. Intelligent social animals like dogs and monkeys seem to be capable of forming alliances, and even “friendships”. Less intelligent, non-social animals, may not be. I have known several rabbits and guinea pigs, and never got the impression that they had a great deal of affection for anyone. Parrots, on the other hand, often seem to like some people very much, and others not at all (and make very clear which is which).

For most conscious animals, the best guess is that consciousness is a raw, moment to moment awareness, consisting mostly of current perceptions, sensations like pain, pleasure and

hunger, and emotions like anger or disgust. Consciousness in such animals is about what is happening in the here and now. Not much time is likely to be spent dwelling on the past, on things not currently present, or on imagining the future. A few very intelligent animals, like bears, monkeys, ravens, or parrots, may be able let their imagination roam beyond the here and now, visualizing situations that might come up at a later date. But humans are the world's masters at letting our minds wander. Whereas for most animals, the play in the theater of consciousness concerns what is happening right now, we have learned to perform exotic plays about the far future, about distant worlds, imagined objects, and abstract philosophical questions.

### **FURTHER ELABORATIONS: BEYOND TRIAL AND ERROR**

Our extensive elaboration on the basic themes of consciousness is perhaps what sets us apart more than anything else. To think about how many possibilities it opens up, it helps to think back on how nervous systems and minds first evolved, and the purposes they originally served. I began this section of the book discussing the functions of a nervous system. One function is helping the genome orchestrate the workings of a complex body. Another is to develop behavioral adaptations that can occur within an organism's lifetime, instead of over many generations. That type of adaptation is known as learning.

Learning might be called a triumph for the individual, because it allows individual animals to learn new tricks for themselves, instead of letting evolution happen upon them long after they are dead. But this very individualism is the big failure of simple learning. For most animals, learning doesn't pass from individual to individual, or build up over the generations like genetic adaptations do. What most animals learn dies with them, because they have no way of passing it on to others, who have to start from scratch. This is why, for most simple animals, learning never got very sophisticated. Letting animals learn by trial and error every generation is not worthwhile, because trial and error is so risky. Tasting the wrong plant can be fatal. This is why, for simple animals from sea snails to lizards, learning is a simple affair, and most behavior is hard-wired. The only time it makes sense to rely on learning a great deal is in animals with very complex, unpredictable lifestyles or environments, which are too variable for canned responses. Elaborate nervous systems and big brains are expensive things, using a great deal of energy. Add

that to the risks of trial and error learning, and you can be sure that most animals will have no more brainpower than they really need. A grasshopper has no use for a big, expensive brain, so it doesn't have one.

## **EXPLICIT THINKING: VIRTUAL TEST RUNS**

Of course, there are ways of going beyond risky trial and error. One is explicit thought—creating and manipulating mental models of the world. It's much safer to imagine what might happen if you did something rash than actually doing it, and finding out what happens the hard way. Prehistoric people who could think “What's the best way to kill that wildebeest? Better not try it alone..” lived much longer than those who just ran at it and hoped for the best. If you can create a model world in your head, and perform test runs there, you can make a more considered judgement about how to proceed. As I have said, some very smart animals are probably capable of a simple version of this, relying entirely on internal sensory imagery. But we humans are really good at it, partly because we can also think with internal dialogue, abstract notions like truth and beauty, and culturally transmitted ideas and images. And that brings us to the second way of going beyond trial and error learning.

## **LEARNING FROM OTHERS: THE BEGINNING OF CULTURE**

Another way to reduce the risk of trial and error is to learn from others. An animal that can learn from others how to find food or avoid predators has a large advantage over one that relies on trial and error. It doesn't run such a risk of making mistakes that could kill it. What this requires is some way to transmit knowledge from one animal to another, such as imitation or language. Whatever the medium, when creatures learn from one another, the speed and efficiency of learning goes into a higher gear. They can learn something new and pass it on to others, instead of letting them learn it for themselves. If knowledge is transmitted from generation to generation, it can start to accumulate. Now there is a body of knowledge that grows over the generations, which may help succeeding generations get along better than previous ones. This is a whole new way of getting along in the world, faster than biological evolution and less risky than

individual learning. This is culture.

Cultural transmission and the ability to explicitly imagine hypothetical situations are powerful tricks, and we humans were not the first to discover them. Chimpanzees, for example, are capable of both. They can solve simple problems in their heads, as well as teach others to do things like fish for termites with a branch. Many animals are capable of simple cultural transmission. Birds learn to sing in regional dialects. Monkeys have been observed learning from each other how to wash their food. But we are the animals who have gotten really good at these things. We have combined them together, along with other impressive skills, in an explosive mixture that has allowed us to expand into practically every niche on Earth, and caused us to grow more clever and powerful all the time. The emergence of imagination and culture signals a major transition in the Earth's history, building new levels of sophistication on top of older styles of biological and neural adaptation. Culture and imagination create new worlds, worlds of our own creation that have added to and modified older worlds, for better and for worse. And so we move from the wide natural world to its most elaborate subset—the human world.

